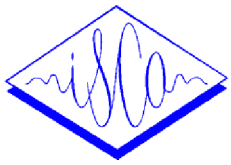
ISCA Archive
http://www.isca-speech.org/archive



Spoken Language Technologies for
Under-Resourced Languages
(SLTU-2010), Penang, Malaysia

INITIALIZING ACOUSTIC PHONE MODELS OF UNDER-RESOURCED LANGUAGES: A CASE-STUDY OF LUXEMBOURGISH

Martine Adda-Decker, Lori Lamel & Natalie D. Snoeren

LIMSI-CNRS, BP 133
91403 Orsay cedex, FRANCE
{madda,lamel,nsnoeren}@limsi.fr

ABSTRACT

The national language of the Grand-Duchy of Luxembourg, Luxembourgish, has often been characterized as one of Europe's under-described and under-resourced languages. In this contribution we report on our ongoing work to take Luxembourgish on board as an e-language : an electronically searchable spoken language. More specifically, we focus on the issue of producing acoustic seed models for Luxembourgish. A phonemic inventory was defined and linked to inventories from major neighboring languages (German, French and English), with the help of the IPA symbol set. Acoustic seed model sets were composed using monolingual German, French or English acoustic model sets and corresponding forced alignment segmentations were compared.

Next a super-set of multilingual acoustic seeds was used putting together the three language-dependent sets. The language-identity of the aligned acoustic models provides information about the overall acoustic adequacy of both the cross-language phonemic correspondances and the acoustic models. Furthermore some information can be gleaned on inter-language distances : the German acoustic models provided the best match with 54.3% of the segments aligned using German seeds, 35.3% using the English ones and only 10.4% using the French acoustic models. Since Luxembourgish is considered a Western Germanic language close to German, this result is in line with its linguistic typology.

Introduction

Luxembourg is a small country in Western Europe, bordered by Belgium, France and Germany (see Figure 1). The national language Luxembourgish ("Lëtzebuergesch") is the language spoken by native Luxembourgers. From a linguistic typological point of view, Luxembourgish belongs to the West central dialects of High German and is therefore part of the Germanic Franconian languages. Just like the English language, Luxembourgish can be considered as a mixed language with strong Romance and Germanic influences. Because of the fact that Luxembourgish is embedded in this multilingual context on the divide between Romance and Germanic cultures, people switch from one language

Fig. 1. Geographical situation of Luxembourg in the heart of Western Europe and on the globe.

to another fairly easily. Therefore, the linguistic situation in Luxembourg poses a real challenge for researchers concerned with both automatic and human language processing for at least two reasons. First, Luxembourgish is strongly embedded in a multilingual context, entailing frequent code-switching and code-mixing. Luxembourgish hence represent an interesting testbed for multilingual processing [1]. Second, Luxembourgish may be considered as a partially under-resourced language, as the written production remains relatively low. Such languages currently represent a hot topic in the field of automatic speech processing, because of a limited written production of Luxembourgish, a poorly observed writing standardization (as compared to other languages such as English and French), and a large diversity of spoken varieties.

In the next section we give some more insight into the linguistic situation in Luxemburg, with a focus on the luxembourgophone situation. Section 2 presents the phonemic inventory of Luxembourgish and its link with other major Western languages. Section 3 presents alignment results with sets of monolingual and multilingual seed models. Section 4 summarizes the achieved results and discusses some major future

challenges for both speech technologies and linguistic studies of Luxembourgish.

1. LINGUISTIC SITUATION OF LUXEMBOURGISH

1.1. Multilingual context

Luxembourg, a small country of less than 500,000 inhabitants in the center of Western Europe, is composed of about 65% of native inhabitants and 35% of immigrants. The national language Luxembourgish is considered the official language of Luxembourg only since 1984. Luxembourgish is the (Moselle Franconian) language spoken by native Luxembourgers, French and German being easily used for communication among residents [2]. Major languages practiced by immigrants used to be Portuguese and Italian. The immigrated population generally speaks or learns one of Luxembourg's other official languages : French or German. Recently, English has joined the set of prestigious languages of communication, and tends to become a major communication tool in professional environments.

Although the country is often considered a successful example of a multilingual society, the linguistic situation of Luxembourg remains complex. Different reasons contribute to this. First, the small size of the country entails a dependance on neighboring countries (Germany, France, Belgium) with a very high rate of cross-boundary exchanges. Moreover, its historical background and its geographical situation puts Luxembourg at the frontier of the Germanic and Romance worlds. Last but not least, an important proportion of immigrants of different linguistic origins adds to the complex linguistic situation that can be observed in Luxembourg.

1.2. An under-resourced language

As was pointed out by [3] and [4], Luxembourgish should be considered as a partially under-resourced language, mainly because of the fact that the written production remains relatively low and that linguistic knowledge and resources, such as lexica and pronunciation dictionaries, are sparse. Rather surprisingly, written Luxembourgish is not systematically taught to children in primary school : German is usually the first written language learned, followed by French [5]. A number of proposals for standardizing the orthography of Luxembourgish can be traced back to the middle of the 19th century. Recently, a successful standardization eventually emerged from the work of a number of specialists charged with the task of creating a dictionary that was published between 1950 and 1977 [6]. The latest spelling reform [2] has been adopted in 1999, and is being used to create official language resources (Cortina, CPLL dictionary). Nonetheless, up until today, German and French are the most practiced languages for written communication and administrative purposes in Luxembourg, guaranteeing a larger dissemination, whereas Luxembourgish is mainly being used for oral communication. It is precisely because of the strong influence of both German and French that Luxembourgish exhibits a large amount of both pronunciation and derived potential writing variants. Pronunciation variants may give rise to resulting variations in written Luxembourgish, as Luxembourgish orthography strives for phonetic accuracy [2]. The question then arises, in particular for oral transcripts, whether the written form reflects the perceived pronunciation form or whether some sort of normalization process is at work that eliminates part of the variation. With respect to automatic speech recognition, text normalization is an important issue in order to achieve reliable estimates for n-gram based language models. In sum, Luxembourgish is predominantly a spoken language that tends to reproduce the observed variations when written.

The limited production of written material is related to the easy use of French and German as written communication languages. Further, no orthographic standards were clearly established before the end of the 20th century. This implies a high degree of variation in the observed written forms. An exhaustive Luxembourgish dictionary was produced after World War II, and this large scale effort actively contributed to the elaboration of spelling standards settled in 1975 and revised in 1999) [7, 8]. Written Luxembourgish sources, although not very widespread, can yet be found over the last decades and even centuries. It is difficult to estimate the numbers of Romance/Germanic influenced words in Luxembourgish, as proportions greatly depend on communicative settings. Nonetheless, one may note that vernacular Luxembourgish is highly influenced by its Germanic filiation, whereas more technical and administrative jargons include a higher proportion of Romance words. Examples in Table 2 are almost all of Germanic influence, except those illustrating nasal vowels, and the /ʒ/and /ɥ/ consonants.

Beyond written material, the existence of sibling resources, providing similar content in both written and audio modalities are particularly helpful for automatic speech recognition (ASR). Steps to an autonomous ASR system include acoustic modeling, the development of a pronunciation dictionary and language modeling [9]. Most languages make use of broadcast news audio data, together with, as written sources, newspaper texts, news wires and related web pages. In Luxembourg news broadcasts are proposed in Luxembourgish on a daily basis, however newspapers remain for the most part bilingual German/French, with only limited code-switching and code-mixing to Luxembourgish, generally for titles. Yet, it is important to highlight recent efforts that have been made regarding the establishment of word lists and multilingual dictionaries in electronic form [10]. Furthermore concerning the WEB, Luxembourgish actually holds rank 55 in the list of 272 official Wikipedias, published by the Wikimedia Foundation for various languages (http ://meta.wikimedia.org/wiki/List_of_Wikipedias). The number of Luxembourgish native speakers can be estimated to 300,000. The immigrated population and the number of daily cross-boarder commuters has steadily increased over the past decades. A relatively high number of more or less proficient

L2 speakers can be found among them, especially as they express a great interest in learning the basics of the Luxembourgish language.

1.3. Luxembourgish corpora

As was mentioned before, sibling resources, providing both audio and related written material are of major interest for ASR development. The most relevant resource we found here, consists in the *Chamber* (House of Parliament) debates and to some extend in news channels, such as delivered by the Luxembourgish radio and television broadcast company RTL.

The Parliament debates are broadcast and made available on the official web site (www.chd.lu), together with written reports (the *Chamber* reports), which correspond to rather close manual transcripts of the oral debates. Another interesting sibling resource stems from the Luxembourgish radio and television broadcast company RTL, which produces news written in Luxembourgish on its web site (www.rtl.lu), together with the corresponding audio data. However only very limited amounts of written Luxembourgish can be found here, whereas RTL has a profuse audio/video production. Table 1 summarizes the different text and audio resources currently being collected. 12M words have been extracted from the

Table 1. Major Luxembourgish text and audio sources for ASR. Collected amounts are given in number of words

	written	**sibling : audio+written**	
Source :	WIKIPEDIA lb.wikipedia.org	CHAMBER www.chd.lu	RTL www.rtl.lu
Volume :	500k	12M/(300h)	700k/(40h)
Years	2008	2002-2008	2007-2008

Chamber reports (years 2002-2008), which mainly comprise professionally transcribed oral debates. However they also include some written subjects in French. The collected audio data correspond to the debates of the two most recent years, totalling a volume of approximately two hundred hours.

2. PHONEMIC INVENTORY

The word lists derived from the written material allow to fix optimal vocabularies for the ASR system. A further step consists in providing pronunciations for each lexical entry. Such pronunciations rely on a phonemic inventory. Hereafter we will give details about the Luxembourgish phonemic inventory, detailing vowels, diphthongs and consonants.

The adopted Luxembourgish phonemic inventory includes a total of 60 phonemic symbols plus 3 extra-phonemic symbols (for silence, breath and hesitations). Table 2 present the selected phonemic inventory together with illustrating examples. Luxembourgish is characterized by a particularly high number of diphthongs. To minimize the phonemic inventory size, we could have chosen to code diphthongs using two consecutive symbols, one for the nucleus and one for the offglide (e.g. the sequence /a/ and /j/ for diphthong a͡ɪ). We prefered the option of coding diphthongs and affricates using specific unique symbols. Given the importance of French imports, nasal vowels, although not required for typical Luxembourgish words, were included into the inventory. Furthermore, the native Luxembourgish makes use of a rather complex set of voiced/unvoiced fricatives.

Concerning linguistic studies [11], many aspects of the Luxembourgish language have been explored on limited spoken material. They still need to be investigated on a larger scale and on fluent speech, in particular for pronunciation variants. The existing phonetic, phonological, prosodic, lexical and morphosyntactic studies are generally carried out using limited objective observations. Large oral corpus-based studies might be carried out, provided Luxembourgish automatic speech alignment and transcription systems were available.

In the following, we raise some issues concerning high-quality pronunciation dictionaries.

2.1. Spelling

Luxembourgish spelling standards aim at minimizing pronunciation ambiguities, even though minor problems remain. For example, the `au` letter sequence is ambiguous with respect to /ɛʊ/ (`Haut`) or /ʌʊ/ (`haut`) pronunciations.

Concerning Romance or Germanic origins of Luxembourgish lexical entries, writing standards may stay more or less close to the language of origin, as discussed in section 1.1. For French words such as `attaquer` (eng. to attack) or `abdiquer` (eng. to abdicate), the corresponding lëtzebuergesch orthographic forms are `attackéieren` and `abdiquéieren` (after the official Luxembourgish CORTINA spellchecker [1]). For Romance items, different pronunciation rule sets need to be developed, that differ from Germanic or Moselle-Franconian pronunciation rules. For instance, depending on the origin, `qu` letter sequence of germanic items such as `quälen, quëtschen, Quetschen` calls for a /kw/ pronunciation, whereas Romance rules generally demand a simple /k/ pronunciation.

3. ALIGNMENT EXPERIMENTS

Alignment experiments are carried out using different initializations for the Luxembourgish acoustic models and different pronunciation dictionaries.

3.1. Acoustic seed models

Many researches have addressed the issue of building acoustic seed models for underresourced languages [12].

[1] More information about the Cortina Luxembourgish spellchecker can be found at http ://cortina.lippmann.lu.

In this work three sets of context-independent and gender-independent acoustic models were built, one for each seed language (i.e., English, French, German). The models were trained on manually transcribed audio data (between 40 and 150 hours) from a variety of sources, using language specific phone sets. The amount of data used to train the acoustic models and the number of phonemes per language are given in Table 3. Each phone model is a tied-state left-to-right, 3-state CDHMM with Gaussian mixture observation densities (typically 32 components). The acoustic features are derived from

Table 3. Characteristics of English, French and German original acoustic model sets.

Language	*#phonemes*	*#training (h)*
English	48	150
French	37	150
German	49	40

a PLP-like [13] acoustic parametrization, which has been used in the LIMSI systems since 1996. The speech features consist of 39 cepstral parameters derived from a Mel frequency spectrum estimated on the 0-8kHz band every 10ms. For each 30ms frame, the Mel scale power spectrum was computed, and the cubic root taken, followed by an inverse Fourier transform. LPC-based cepstrum coefficients were then computed. These cepstral coefficients were normalized on a segment cluster basis using cepstral mean removal and variance normalization. Each resulting cepstral coefficient for each cluster has a zero mean and unity variance. The 39-component acoustic feature vector consists of 12 cepstrum coefficients and the log energy, along with the first and second order derivatives.

Four sets of pseudo-Luxembourgish acoustic models, each including 63 phones, were created from the English, French and German seed models by mapping the Luxembourgish phonemes to a close equivalent in each of the three model sets. Table 2 include the adopted cross-lingual associations, to initialze seed models for Luxembourgish. It can be noted that some symbols are used several times for different Luxembourgish phonemes. In particular, for the diphthongs, which are missing in French, we chose to select the phonemes corresponding to the nucleus vowel. A fourth model set was constructed by concatentating the first three model sets, so that the decoder could chose among the three languages' models (see Table 4). For each word, choose the acoustic models from the language with the best match.

3.2. Multilingual pronunciation variants

For the alignment experiments using the language-dependent phone sets the same pronunciation dictionary was used. We introduced some variants for the most frequent function words, French imports and some variants to account for word-final mobile-n deletion (or insertion) [4]. Example variants are shown in Table 5.

Table 4. Pseudo-Luxembourgish acoustic models using either English, French and German acoustic model sets or a super-set of multilingual acoustic seeds.

Language	*#phonemes*	*#training (h)*
English	63	150
French	63	150
German	63	40
Super-set (E,F,G)	3x63	340

Table 5. Excerpt of the Luxembourgish pronunciation dictionary as used for the proposed alignments. The upper part shows typical examples of variants (frequent words, French loan words, mobile-n deletion). The lower part illustrates the pronunciation dictionary used for alignments with the multilingual acoustic super-set.

lexical entry (English)	*citation form*	variants
déi (those)	d͡eɪ	dɪ
President (president)	pʁəzidɛnt	pʁezidã
Europa	ɔ͡ɪʁopa	øʁopa
an (and, in)	ʌn	ʌ
Multilingual dictionary		
déi (those)	d_g e͡ɪ$_g$ d_f e͡ɪ$_f$ d_e e͡ɪ$_e$	d_g ɪ$_g$ d_f ɪ$_f$ d_e ɪ$_e$

3.3. Luxembourgish audio alignment

The Luxembourgish audio corpus with corresponding detailed acoustic transcripts comprised 80 minutes of hand transcribed audio data (Chamber (70') and News (10')). We produced these detailed transcripts from scratch for the news data. For the Chamber data, the audio stream was manually segmented into speaker turns, according to the existing *bona fide* report. For each speaker, the *bona fide* transcriptions were changed if necessary to faithfully reflect the speech flow. All uttered audible speech events, including disfluencies and speech errors were manually transcribed. The quality of the manual *verbatim* transcripts were checked via the resulting word lists for typos and orthographic inconsistencies. The transcript quality further needs to be questioned, if significant amounts of data are rejected during alignment. As the same transcripts were used for the different Luxembourgish acoustic seed models, if more data are rejected for a given model set than for the others, this set may be considered as less appropriate, without blaming the transcripts.

The percentage of the audio data aligned with phone segments varies from 77-80%, the lowest figure corresponding

Table 6. Total duration (in seconds) aligned as phones, as extra phonemic segments (silence, breath or hesitation) or rejected due to model/data mismatch.

Language	*phon.dur.*	*#extra dur.*	rejected
English	3910	673	516
French	3933	790	373
German	4043	921	131
Super-set (E,F,G)	4077	814	203

to English, the highest to the multilingual and German configurations. The remaining 20-23% of the acoustic data are either aligned with extra-phonemic symbols or rejected by the alignment system, due to model/data mismatch. It can be noted that English has the highest rate of rejected data : 516 seconds which correspond to 10% of the data. Such a high rejection rate normally would require to check the manual transcripts and/or the pronunciation dictionary. Fortunately, for the other configurations, the rejection rates are much lower, the lowest rates being achieved by the German language (131 seconds, $< 3\%$). German has the highest contribution to the extra-phonemic symbol set.

The average phone segment duration remains almost stable with respect to the different monolingual seed alignments. Variations here stem from variable proportions of the acoustic signal assigned to the extra-phonemic models. The German alignment yields the smallest average phone duration of 0.07 seconds on average (silence, breath and hesitation segments are not considered). For English and French the average segment duration corresponds to 0.08 seconds. We could observe that independently of acoustic-phonetic considerations, the (German) silence (including background noise) model was made use of more frequently during the German monolingual alignment, than was the case for the French or English silence models. This explains the smaller average phone duration. This might be related to the relatively small volume of training data (40h) for the German originated seeds (as opposed to French and English), with a lower capacity to cover various acoustic conditions.

The results presented in Table 6 further suggest that the German acoustic models are globally best at explaining the Luxembourgish data, as the smallest volume of data was rejected.

On a more linguistic level of analysis, the results show that unvoiced segments tend to be longer than their voiced counterparts, and that diphthongs and nasal vowels are about 30% longer than oral vowels. More precise results on the Luxembourgish phonemes will be produced in the future, with acoustic models trained on a larger set of Luxembourgish data.

3.4. Multilingual alignments

The alignment produced by the acoustic super-set model, together with the multilingual pronunciation dictionary achieves the highest proportion of aligned acoustic phone segments. In this configuration, it is interesting to investigate the results on two levels : (i) on the phone segment level, we can measure the proportions of segments aligned using the seeds of a given language. Are there differences in proportions as a function of phonemes ? (ii) on the word level, we may check whether the proportion of aligned French seeds is higher for French loan words than for native Luxembourgish words.

For example, we may expect that for Luxembourgish diphthong segments, the proportion of aligned English seeds may increase, especially for diphthongs not covered by the German language. Conversely the proportion of French and English seeds used for Luxembourgish and German specific sounds (e.g. χ) should remain very low.

Table 7 displays aligned monolingual seed proportions as produced by the multilingual super-set. More than half of the 55873 segments were aligned using the German seeds. About one third corresponds to English seed models and only 10% of the segments were aligned using the French models. Results for some phonemes are shown to illustrate that proportions can notably vary with phoneme identity.

Table 7. Proportions of aligned German, English, French seeds in the multilingual super-set configuration. The number of phone occurrences is provided. Results are given on average and a subset of selected phonemes.

Phone type	*German*	*English*	*French*	# occ.
overall	54.3	35.3	10.4	55873
p	67.05	21.85	11.10	865
t	55.91	35.23	8.86	3588
k	55.15	36.64	8.21	1048
ç	56.80	34.52	8.67	588
χ	80.87	14.29	4.84	413
h	36.05	59.36	4.59	785
ʒ	41.96	25.00	33.04	112
y	25.00	15.62	59.38	32
ʏ	41.03	25.64	33.33	39

4. SUMMARY AND PROSPECTS

The main goal of the present contribution was to draw attention to the complex linguistic situation of Luxembourgish, a partially under-resourced and under-described language. For ASR development, the use of sibling resources that provide similar contents in both written and oral/auditory modalities is extremely useful. Although there are relatively few written resources in Luxembourgish as compared to other European languages, corpus studies in Luxembourgish will

substantially add to the current debate on the processing of pronunciation variants in automatic and natural speech processing.

In the present work, we focused on the issue of producing acoustic seed models for Luxembourgish. A phonemic inventory was defined and linked to inventories from major neighboring languages (German, French and English), with the help of the IPA symbol set. For each of these languages, acoustic seed models were composed using either monolingual German, French or English acoustic model sets. During Luxembourgish speech alignments, a super-set of multilingual acoustic seeds was used putting together the three language-dependent sets. The language-identity of the aligned acoustic models provides information about the overall acoustic adequacy of both the cross-language phonemic correspondances and the acoustic models. Furthermore some information can be gleaned on inter-language distances : the German acoustic models provided the best match with 54.3% of the segments aligned using German seeds, 35.3% using the English ones and only 10.4% using the French acoustic models. Since Luxembourgish is considered a Western Germanic language close to German, this result is in line with its linguistic typology.

Computational ASR investigations and corpus-based analyses will not only enhance the development of a more full-fledged ASR system for Luxembourgish, but can also be used to generate more specific predictions about the role of the actual experience that listeners have with pronunciation variants. In turn their predictions can then be tested in other domains such as psycholinguistics. Given the implications of large corpus-based analyses, it is hoped that this line of research on Luxembourgish will sparkle more interest for this language in researchers working in the domains of ASR and linguistics.

Acknowledgements

The authors would like to thank the members of the CPLL (Permanent Council of the Luxembourgish Language) for their help with sibling corpora and fruitful discussions. This work has been partially financed by OSEO under the QUAERO program. Part of the research has been supported by an FNR grant from the Luxembourgish gouvernment awarded to N.D.S.

5. REFERENCES

[1] M. Adda-Decker and L. Lamel, *Multilingual pronunciation dictionaries in Multilingual Speech Processing*, Elsevier, 2006.

[2] F. Schanen, *Parlons Luxembourgeois*, L'Harmattan, 2004.

[3] M. Adda-Decker, T. Pellegrini, E. Bilinski, and G. Adda, "Developments of letzebuergesch resources for automatic speech processing and linguistic studies.," in *LREC*, 2008.

[4] C. Krummes, "Sinn si or si si ? mobile-n deletion in luxembourgish," in *Papers in Linguistics from the University of Manchester : Proceedings of the 15th Postgraduate Conference in Linguistics*, Manchester, 2006.

[5] C. Berg and C. Weis, "Sociologie de l'enseignement des langues dans un environnement multilingue. rapport national en vue de l'élaboration du profil des politiques linguistiques éducatives luxembourgeoises," Tech. Rep., 2005.

[6] P. Linden, *Luxemburger Wörterbuch*, P. Linden, Hofbuchdrucker, 1950.

[7] G. Newton, *Studies from the Germanic Languages - The standardization of Luxembourgish*, John Benjamins Publishing Company, 2002.

[8] F. Schanen and J. Lulling, "Introduction à l'orthographe luxembourgeoise.," in *www.cpll.lu/ortholuxs_l.html*, G.-D. de Luxembourg, 2003.

[9] L. Lamel, J.L. Gauvain, and G. Adda, "Lightly supervised and unsupervised acoustic model training," *Computer Speech and Language*, vol. 16, no. 1, pp. 115–229, 2002.

[10] J. Lulling, *Luxdico - dictionnaire bilingue luxembourgeois - français*, Presses universitaires de Namur, 2005.

[11] C. Moulin, *Perspektiven einer linguistischen Luxemburgistik - Studien zu Diachronie und Synchronie*, Universitätsverlag WINTER Heidelberg, 2005.

[12] T. Schultz and A. Waibel, "Experiments on cross-language acoustic modeling," in *Proceedings of Eurospeech*, Aalborg, 2001.

[13] H. Hermansky, "Perceptual linear predictive (plp) analysis of speech.," vol. 87, no. 4, pp. 1738–1752, 1990.

Table 2. Cross-lingual phone association table. Luxembourgish target phonemes are associated to identical or similar (in grey) phonemes of the different French, German, English source languages.

Carrier word (Eng.)	Lux.	Fre	Ger	Eng
ORAL VOWELS				
l**ii**cht (light)	i	i	i	i
L**i**dd (song)	ɪ	i	ɪ	ɪ
S**ü**den (south)	y	y	y	i
sch**ü**tzen (shelter)	ʏ	y	ʏ	ɪ
L**ee**d (sorrow)	e	e	e	e
z**é**ng (ten)	ɛ	e	e	ɛ
f**ä**eg (able)	ɛ :	ɛ	ɛ :	ɛ
h**ä**tt (would)	ɛ	ɛ	ɛ	ɛ
F**öh**n (hairdryer)	ø	ø	ø	ɔ
m**ë**ll (soft)	œ	œ	œ	ə
et (it)	ə	ə	ə	ə
h**a**t (had)	a	a	a	ɑ
h**a**tt (she)	ʌ	a	ʌ	ɑ
R**o**t (advice)	o	o	o	o
L**o**ft (air)	ɔ	ɔ	ɔ	ɔ
L**uu**cht (lamp)	u	u	u	u
H**u**tt (hat)	ʊ	u	ʊ	ʊ
NASAL VOWELS : French imports				
enf**in**	ɛ̃	ɛ̃	ɛ	æ
enfin	ã	ã	a	ʌ
b**on**	õ	õ	o	o
DIPHTHONGS				
l**ie**wen (to live)	ɪə	i	i	i
l**éi**en (to tell lies)	eɪ	e	e	e
l**äi**t ((he) lies down)	ɛɪ	e	e	e
l**au**schteren (to listen)	ɛʊ	ɛ	ɛ	æ
l**ei**en (to lie down)	aɪ	a	aɪ	aɪ
l**au**den (to ring)	aʊ	a	aʊ	aʊ
Europa	ɔɪ	ɔ	ɔɪ	ɔɪ
l**ou**nen (to hire)	ɔʊ	o	o	o
l**ue**wen (to praise)	ʊə	u	ʊ	ʊ
SYLLABICS				
Kann**er** (children)	ɐ	ə	ɐ	ʌ
fein**em** (fine)	m̩	m	m̩	m̩
laf**en** (to run)	n̩	n	n̩	n̩
eid**el** (empty)	l̩	l	l̩	l̩

Carrier word (Eng.)	Lux.	Fre	Ger	Eng
PLOSIVES				
paken (to package)	p	p	p	p
taasch**t**en (to touch)	t	t	t	t
kachen (to cook)	k	k	k	k
baken (to bake)	b	b	b	b
droen (to carry)	d	d	d	d
goen (to go)	g	g	g	g
FRICATIVES & AFFRICATES				
Feier (fire)	f	f	f	f
lue**s** (slow)	s	s	s	s
Zuch (train)	ʦ	s	s	s
Schoul (school)	ʃ	ʃ	ʃ	ʃ
Ee**ch**en	ç	ʃ	ç	ʃ
Zu**ch** (train)	χ	k	χ	k
Hand (hand)	h	{br}	h	h
Wieder (weather)	v	v	v	v
Summer (summer)	z	z	z	z
Gilet (vest)	ʒ	ʒ	ʒ	ʒ
Lig**e**n (lie)	j	ʒ	ç	ʒ
NASALS & GLIDES				
Mamm (mother)	m	m	m	m
Noper (neighbour)	n	n	n	n
m**éng** (mine)	ŋ	n	ŋ	ŋ
Leit (people)	l	l	l	l
Rou (rest)	ʁ	ʁ	ʁ	ɹ
H**ä**r (mister)	ɐ̯	ə	ɐ̯	ə
Suite (suite)	ɥ	ɥ	ʏ	w
Juli (July)	j	j	j	j
Quetsch (plum)	w	w	ʊ	w
EXTRA-PHONEMIC SYMBOLS				
silence	{sil}	{sil}	{sil}	{sil}
hesitation	{hes}	{hes}	{hes}	{hes}
breath	{br}	{br}	{br}	{br}